# Common-Near-Neighbor Information in Discriminative Spaces for Human Re-identification( Digest_要約 )

AUTHOR(S):

Li, Wei

# Common-Near-Neighbor Information in Discriminative Spaces for Human Re-identification

LI WEI

April 2014

# Abstract

Matching people across camera views at different sites, known as human re-identification, is challenging and valuable for both academia and industry. To date, attempts to address this issue have involved feature representation and/or dissimilarity measurement. Although several improvements have been achieved, the problem is still far from being solved because of the real-world complexities. These complexities involve large within-class variations due to the changeable body appearance and environment and small between-class differences arising from the similar body shape and clothing style.

In the literature, there are many works focusing on feature representation. However, to build an accurate correspondence between highly variable elements (sample points or sample sets) of human image data in the feature space, a reliable dissimilarity measurement is indispensable. Conventional dissimilarity directly describes how far apart the concerned pair of elements they are. Although this kind of dissimilarity has been generalized in a variety of ways in the subsequent development, still it solely relies on the pair of measured elements without considering the neighborhood information. And for re-identification, this kind of dissimilarity measurement seems to confront a dilemma to discriminate the isolated elements that have large within-class variations and small between-class differences.

In the thesis, we reformulate re-identification as the problem of finding the correct matches between the elements from the query side and from the corpus side based on a reliable dissimilarity measure. Motivated by the idea to deliver the effectiveness of those well-distributed elements to those badly-distributed elements in a metric space, we creatively propose to quantify the local neighborhood structure of the pair of elements in each other's neighborhood structures into the dissimilarity, as a new trial to enrich the conventional distance notion by considering the neighborhood information. Here, neighborhood structure is defined as the layout relationship between the concerned element and its neighbors in the given met-

ric space. The property of this dissimilarity can be comprehended by the insight of measuring the quantity of common nearness for the pair of elements in each other's neighborhood structure, which we refer to as "common-near-neighbor information". To analyze the common-near-neighbor information, a discriminative metric space is indispensable. Such metric space can be constructed and improved in consideration of intra-class compactness as well as inter-class separation.

Revolving around the philosophy of exploiting the common-near-neighbor information in a discriminative metric space, we have presented various approaches for the re-identification problem in both single-shot and multiple-shot directions. Single-shot and multiple-shot cases have different prerequisites and evaluation principles. For the single-shot case, we need to match each query point to each corpus point one after another, while for the multiple-shot case, we are able to simultaneously decide the matching between the whole set of query points and the whole set of corpus points. Single-shot and multiple-shot re-identification cases own distinct resources and challenges. It is these resources and challenges that necessitate the divide-and-solve strategy for exploiting the common-near-neighbor information. For instance, the point based distance seems more manageable, due to the relatively simple information for each identity in the single-shot case, than the set based distance that is affected by the local within-set variations in the multiple-shot case, while adequacy of the within-set distribution information enhances the reliability of the set based distance in the multiple-shot case instead.

In greater details, against the single-shot re-identification case, not only has the capability of point-level common-near-neighbor information for classifying the complex human image data been confirmed, but also the importance of the metric space where the point-level common-near-neighbor information is exploited has been studied, as elaborated in Chapter 2. Indeed, by making intra-class distances smaller than inter-class distances between sample points, a suitable learned Mahalanobis metric space can benefit modeling the point-level common-near-neighbor information into the dissimilarity. During the specific single-shot vs. single-shot re-identification procedure, to compensate for the insufficiency of the within-class distribution information, we further explore the strengthened metric space by two approaches for the subsequent point-level common-near-neighbor information exploitation, as expounded in Chapter 3. The first approach couples two complementary metric learning schemes together, and the second approach incorporates the constraints of point-level common-near-neighbor modeling dissimilarity comparison into the metric learning framework.

To resolve the multiple-shot re-identification problem, an intuitive idea is to exploit the set-level common-near-neighbor information, by treating each set of points as a whole, without ignorance of the set based integrity and within-set distribution, as suggested in chapter 4. For effectively measuring this dissimilarity, the metric spaces are selectively constructed by two kinds of representative features. One feature addresses condensing the sets into the representative imaginary points based on covariance descriptors in the Riemannian space. The other feature emphasizes compacting human image sets to depress the outliers and intruders by collaborative representation in the Hilbert space. Though discriminative, these feature spaces are expensive in fact. Representative covariance descriptor requires adequate and typical within-set images. Discriminatory collaborative representation requires diverse and labeled dictionary data. We expect to avoid these expensive requirements for the single-set vs. single-set case, and designs a method based on mining the locality information, as detailed in Chapter 5. In this method, a capable set-to-set distance is crafted by encoding the local minority distribution information between paired sets, and upon this distance, an effective metric field space is constructed to accommodate the local variation of each set, before the set-level common-near-neighbor modeling dissimilarity is measured among all sets.

In addition to theoretic analysis, experimentation across several widely-used benchmark datasets for real-scenario human re-identification has demonstrated not only the philosophical value but also the methodological superiority in the thesis. Future work will cover, but not limited to, developing new models based on this philosophy and applying re-identification to improving cross-camera tracking.